Изосимов Владимир Юрьевич
зав. отделом мониторинга и оценки организаций в сфере науки и инноваций РИЭПП.
Тел. 917-86-66,
info@riep.ru

Напреенко Владислав Георгиевич
кандидат технических наук,
ведущий научный сотрудник
сектора мониторинга состояния и тенденций развития организаций сферы науки и инноваций РИЭПП.
Тел. 917-86-66,
info@riep.ru

К ПРОБЛЕМЕ КЛАСТЕРИЗАЦИИ НАУЧНЫХ ОРГАНИЗАЦИЙ

Решение задач государственного регулирования сферы исследований и разработок – таких как мониторинг состояния, основных тенденций и направлений развития научно-технической сферы в целом и государственного сектора в частности; вовлечение научных организаций в решение задач государственной научно-технической политики; обеспечение координации научных исследований и разработок; размещение государственного заказа и других одновременно – требует проведения классификации (типологизации) множества организаций научно-технической сферы по тем или иным признакам, определяемым реализуемыми целями и задачами. Другими словами, решение конкретных задач реализации государственной политики в сфере исследований и разработок требует дифференцированного подхода к различным типам научных организаций, т. е. требует решения задачи кластеризации научных организаций. Суть процесса кластеризации составляет, как известно, разделение исходной совокупности объектов (в данном контексте – множества научных организаций) на группы наиболее близких по некоторым формальным признакам. В инструментальном плане кластерный метод представляет собой область многомерного статистического анализа, основанного на упорядочении набора количественных и/или качественных данных.

Безусловно, при любом выделении типов научных организаций в возможности их кластеризации возникают обоснованные сомнения. «Насколько принятая типология научных организаций соответствует тем или иным аналитическим задачам? Сколько типов существует реально? Какова естественная типология научных организаций? Если такая типология излишне сложна, не удастся ли ее упростить, рассматривая научные организации с определенных практических позиций – с точки зрения научного уровня, отдачи, финансирования, перспективности и выживания и т. д.» [1].

В данной работе не ставилась задача предложить универсальную типологию научных организаций (если таковая вообще возможна). Более естественным является предположение, что при решении каждой конкретной задачи государственного регулирования сферы исследований и разработок кластеризация научных организаций будет уникальна как по количеству, так и по составу кластеров. Сколько задач – столько и группировок научных

организаций. Поскольку «кластеризация представляет собой разбиение множества на кластеры – подмножества, параметры которых заранее неизвестны» [2]. Кроме того, даже при решении конкретной задачи подходы могут различаться. Нельзя не согласиться с утверждением, что «существует столько способов деления мира на объектные системы, сколько ученых принимается за эту задачу» [3].

Цель данной работы – предложить методику кластеризации научных организаций, применимую при решении широкого круга задач государственного регулирования сферы исследований и разработок.

Предлагаемая методика выделения кластеров научных организаций предполагает пять основных этапов. На первом этапе осуществляется предварительный отбор организаций сферы исследований и разработок, подлежащих анализу в рамках реализации конкретной цели кластеризации. Например, если целью кластеризации является выявление неэффективных научных организаций государственного сектора для их последующей реорганизации, то в анализируемое с этой целью множество научных организаций нет смысла включать те, реорганизация или ликвидация которых невозможна по соображениям государственной безопасности, национального престижа и т. п.

Второй этап заключается в формировании набора количественных и/или качественных данных (системы показателей) для выделения кластеров научных организаций.

В настоящее время в отечественной и зарубежной практике используется более тысячи показателей, описывающих научные организации. Наиболее характерными для отечественных работ можно считать показатели, вошедшие в статистические формы «1-наука» и «2-наука», а также применяемые в бухгалтерской отчетности. Реже используются элементы наукометрии и библиометрии, включая так называемое руководство Фраскати («Manual Frascati» – от названия города в Италии, где в 1963 г. была принята первая версия документа; последняя редакция была принята в 1993 г.) и показатели публикации статей и индексы цитирования в области естественных и общественных наук (например, американские Science Citation Index, Social Sciences Citation Index соответственно). В российских условиях возможности использования индексов цитирования ограничены, что объясняется отсутствием организаций, ведущих полноценную и постоянную работу по сбору, анализу и обобщению соответствующей информации, а также недостаточной информативностью таких индексов в отношении большого числа российских работ.

Множественность существующих показателей делает необходимым их отбор в соответствии с конкретными целями кластеризации. Чтобы осуществить такой отбор, целесообразно прежде всего определить базовые направления кластеризации. Каждое направление можно рассматривать как типовой принцип отбора и группировки показателей. В зависимости от цели кластеризации, при ее осуществлении могут быть использованы разные базовые направления или их сочетания. Тем не менее выбор базовых направлений можно сделать в достаточной мере универсальным, что позволит упростить и унифицировать процесс кластеризации научных организа-

ций, а также сформировать типовые перечни показателей, используемых для решения типовых задач.

Определение базовых направлений кластеризации должно осуществляться с соблюдением следующих требований: определенность, полнота, компактность и ограниченность пересечений направлений по показателям.

Требование определенности в данном случае означает, что каждому из базовых направлений кластеризации должен соответствовать достаточно устойчивый набор показателей. Полнота базовых направлений предполагает возможность отнесения любого из показателей как минимум к одному из направлений. Компактность базовых направлений кластеризации понимается как ограниченность числа типовых показателей, относящихся к каждому направлению. Требование ограниченности пересечений базовых направлений кластеризации означает, что в списке типовых показателей каждого направления не должно быть слишком большой доли показателей, относящихся одновременно к другим направлениям.

Анализ возможных целей кластеризации, а также традиции, сложившиеся в отечественной практике наукометрии, позволяют выделить следующие базовые направления кластеризации научных организаций.

1. Организационно-правовое положение и связи организации.
2. Кадровый и научно-технический потенциал.
3. Научная, инновационная и образовательная деятельность.
4. Экономика организации.

После того как базовые направлениям кластеризации определены, можно перейти к формированию списков показателей, отвечающих каждому из этих направлений. Процесс определения таких списков складывается из следующих операций:

– формируется перечень возможных показателей, из числа которых должны быть выбраны наилучшие;
– по всем показателям определяются источники данных;
– оценивается информативность показателей по каждому базовому направлению;
– выявляются группы показателей, в рамках которых одни показатели могут быть прямо либо косвенно выражены через другие; прямым выражением в данном случае считается наличие точных расчетных формул, связывающих показатели друг с другом; косвенное выражение осуществляется посредством статистических (корреляционных) связей между показателями;
– показатели, не обеспеченные доступными источниками данных, исключаются из списка;
– показатели, обладающие низкой и нулевой информативностью по всем базовым направлениям кластеризации также исключаются.

При этом в группах показателей, связанных корреляционно либо функционально (с одинаковым вектором влияния), к наилучшим показателям относятся те, которые полнее обеспечены доступными источниками данных, а при равной доступности данных – наиболее информативны.

Показатели, характеризующие организационно-правовое положение и связи научной организации в общем случае описывают обозначенные ниже данные:

1) «координаты» организации в системе государственного учета;
2) взаимодействие научной организации с государством и деловыми партнерами;
3) факторы, препятствующие изменению организационно-правового положения и связей научной организации, и планы таких изменений.

К типовым показателям, характеризующим «координаты» организации в системе государственного учета, можно отнести:

- организационно-правовую форму;
- форму собственности;
- вид экономической деятельности;
- вид научной деятельности;
- территориальное положение;
- подчиненность органу государственного управления;
- время создания.

Типовыми показателями, описывающими взаимодействие организации с государством и деловыми партнерами, являются:

- объем государственного финансирования работ;
- объем работ и услуг, выполняемых для реализации полномочий федеральных органов исполнительной власти;
- объем работ по государственным и международным программам;
- объем зарубежного финансирования;
- объем коммерческого финансирования;
- включенность в научно-производственные интегрированные структуры и в элементы инновационной структуры (центры трансфера технологий, технопарки и т. д.).

Примерами факторов, препятствующих изменению организационно-правового положения и связей научной организации, могут служить:

- уникальность организации в качестве основного производителя особого вида продукции (работ, услуг);
- выполнение функций, которые в соответствии с законодательством могут осуществляться исключительно государственными организациями.

Примерами планов по изменению организационно-правового положения и связей научной организации являются планы ведомств по структурному и/или организационному реформированию в целях повышения инвестиционной привлекательности.

К *показателям*, *характеризующим кадровый и научно-технический потенциал*, относятся главным образом показатели кадрового потенциала научной организации и ее имущественного комплекса.

Типовыми показателями кадрового потенциала можно считать:

- общее число работников (среднегодовое);
- число совместителей и работавших по договорам гражданско-правового характера;
- число работников, выполняющих исследования и разработки;

- долю совместителей и работавших по договорам гражданско-правового характера в общей численности работников, выполняющих исследования и разработки;
- долю кандидатов и докторов наук в численности исследователей и разработчиков (без совместителей);
- долю исследователей в численности исследователей и разработчиков (без совместителей);
- число исследователей;
- долю исследователей в возрасте до 30 лет в общем числе исследователей;
- число исследователей и разработчиков (без совместителей), принятых на работу за последний год;
- число исследователей и разработчиков (без совместителей), уволенных по собственному желанию за последний год.

К типовым показателям имущественного комплекса организации допустимо отнести:

- стоимость имущественного комплекса;
- долю центров коллективного пользования и иных элементов инновационной инфраструктуры федерального значения в стоимости имущественного комплекса;
- долю уникальных комплексов (оборудования, установок, стендов, биологических коллекций и т. п.) в стоимости имущественного комплекса;
- долю зданий, участков земли и объектов природопользования в стоимости имущественного комплекса;
- долю нематериальных активов в стоимости имущественного комплекса;
- износ имущественного комплекса.

Показатели научной, инновационной и образовательной деятельности отражают:

1) содержание работ научной организации;
2) инновационную и научную активность;
3) образовательную деятельность.

Типовыми показателями содержания работ научной организации являются:

- тематика работ;
- распределение текущих затрат по тематике работ;
- доля фундаментальных исследований в текущих затратах организации;
- доля прикладных исследований в текущих затратах;
- доля разработок в текущих затратах;
- доля приоритетных направлений работ в текущих затратах.

Типовыми показателями инновационной и научной активности организации можно считать:

- объем работ по внутренним затратам;

- производительность научного труда (внутренние затраты на исследования и разработки к численности исследователей и разработчиков);
- объем инновационных работ;
- наличие у организации сертификата ISO;
- количество полученных свидетельств об открытиях за последние 5 лет;
- количество полученных патентов на изобретения, промышленные образцы и селекционные достижения, а также свидетельства на полезные модели за последние 3 года;
- количество полученных зарубежных патентов за последние 3 года;
- количество реализованных сделок по проданным лицензиям за последние 3 года;
- количество реализованных сделок по уступке патентов за последние 3 года;
- количество правительственных наград и наград международных организаций, полученных организацией за последние 3 года;
- количество публикаций за последние 3 года;
- индекс цитирования публикаций.

Типовыми показателями образовательной деятельности научной организации служат:

- среднегодовое число студентов, одновременно проходящих обучение (стажировку) в организации;
- среднегодовое число аспирантов, одновременно проходящих обучение (стажировку) в организации;
- среднегодовое число соискателей ученой степени кандидата наук, одновременно проходящих обучение (стажировку) в организации;
- среднегодовое число соискателей ученой степени доктора наук, одновременно проходящих обучение (стажировку) в организации.

Типовыми *показателями*, *характеризующими экономические результаты деятельности организации*, являются:

1) источники финансирования работ и размеры финансирования из разных источников (собственные средства, средства бюджета, средства федерального бюджета по приоритетным направлениям, средства сектора высшего образования, средства коммерческих организаций, средства иностранных источников);

2) затраты на проводимые работы с разделением на фундаментальные исследования, прикладные исследования, разработки, а также с разделением по тематике работ;

3) показатели прибыльности и экономической эффективности (прибыль всего, прибыль от научной деятельности, доля оплаты труда в составе внутренних затрат организации, доля оплаты труда исследователей и разработчиков в составе внутренних затрат организации);

4) бухгалтерский баланс организации;

5) показатели финансовой устойчивости, особенно:

- коэффициент дебиторской задолженности (дебиторская задолженность к внутренним затратам);

- коэффициент ликвидности (оборотные активы к краткосрочным обязательствам);
- коэффициент автономии (капитал и резервы организации к пассивам);
- коэффициент кредиторской задолженности (кредиторская задолженность к внутренним затратам).

Окончательный список рассматриваемых показателей определяется с учетом конкретной цели кластеризации. При этом может потребоваться введение дополнительных показателей. Например, учет показателей, характеризующих не столько научную организацию, сколько влияние выполняемых ею работ (например, разработок новых технологий) для экономики страны, социальной сферы, экологии и др. Достаточно важным является также учет риска отдельных наиболее крупных работ.

Влияние работ научной организации на экономику страны учитывает:

- экономический эффект отдельных работ для исполнителей, заказчиков и бюджета;
- возникающие в результате работ изменения научно-технического потенциала России;
- обусловленное результатом работ изменение рыночной ситуации.

Влияние работ научной организации на социальную сферу отражает:

- трудоустройство либо высвобождение работающего населения;
- изменения условий труда и быта;
- миграционные процессы;
- культурно-образовательные изменения.

Экологический эффект учитывает:

- сбережение и потребление невосполнимых ресурсов;
- ценность сберегаемых (потребляемых) невосполнимых ресурсов;
- ликвидацию либо создание загрязнений;
- тяжесть соответствующих загрязнений.

Учет риска может вестись по отдельным направлениям (риск политический, юридический, организационный, экономический, технологический и др.), причем внутри каждого направления при необходимости могут быть выделены более мелкие составляющие. Например, политический и юридический риски крупных разработок могут быть описаны такими составляющими, как:

- соответствие разработки перспективным интересам России;
- соответствие текущей политике;
- соответствие международному законодательству;
- соответствие законодательству РФ;
- ожидаемая реакция лоббистских групп;
- ожидаемая реакция общественности.

Определив списки учитываемых показателей, можно приступать к выяснению их значений. Главными источниками информации о значениях типовых показателей, перечисленных выше, является статистическая и бухгалтерская отчетность научных организаций. По отдельным пунктам требуется привлечение дополнительной информации, предоставляемой ве-

домствами или самими организациями. В отношении дополнительных показателей важную роль способны играть экспертные оценки.

Следующий (третий) этап – *переход к целочисленным балльным оценкам значений показателей*, предусматривающим небольшое число возможных вариантов оценок. Простейшие случаи – 3-х балльная и 5-ти балльная шкалы. Шкалы оценок разных показателей могут не совпадать.

Часть показателей, описывающих научную организацию, с самого начала имеют балльные значения. Например, балльную форму способны иметь экспертные оценки риска работ.

Другая небольшая часть показателей может иметь малое число вариантов значений. В этом случае перейти к балльным оценкам не составляет труда. Так, показатель «выполнение функций, которые в соответствии с законодательством могут осуществляться исключительно государственными организациями», способен принимать два значения: «выполняет» и «не выполняет». Можно условиться считать оценкой значения «выполняет» 1 балл, а оценкой значения «не выполняет» 2 балла.

Для всех остальных показателей переход к баллам требует специальной процедуры оценивания. Она может быть расчетной либо экспертной.

Расчетный путь предполагает учет статистических данных и вычислительную их обработку. Например, пользуясь статистикой данных научных организаций, можно для некоторого показателя X найти его статистическое среднее Ср(X) и среднеквадратическое отклонение СКО(X). Один из вариантов оценки показателя X по трехбалльной шкале определяется следующим образом:

- если X имеет значение менее Ср(X) – СКО(X), то балльная оценка равна 1,
- если X имеет значение не менее Ср(X) – СКО(X) и не более Ср(X) + СКО(X), то балльная оценка равна 2,
- если X имеет значение более Ср(X) + СКО(X), то балльная оценка равна 3.

При экспертной процедуре оценивания числовые границы диапазонов показателей, отвечающих тем или иным балльным оценкам, устанавливаются экспертами.

Следует учитывать, что для организаций различного рода деятельности могут применяться неодинаковые числовые границы диапазонов показателей, определяющие балльные оценки. Допустим, нас интересует балльная оценка технической оснащенности. Тогда научная организация, осуществляющая исследования в сфере социальных вопросов государственного управления, может быть признана высоко оснащенной, если стоимость ее оборудования (компьютеры, оргтехника) составляет 50 тыс. руб. на сотрудника. В то же время для научной организации, осуществляющей исследования в сфере экспериментальной ядерной физики, ту же стоимость оборудования трудно признать высокой.

В условиях различия границ значений показателей, отвечающих балльным градациям оценок, может быть уместна вспомогательная группировка организаций по принципу сопоставимости показателей. В рамках каждой такой группы (вспомогательного кластера) можно применять одинаковые

границы диапазонов показателей, отвечающие одинаковым балльным оценкам. Для разных групп организаций эти границы будут различаться.

Основное влияние на сопоставимость показателей оказывают такие особенности научных организаций, как: вид деятельности; различный уровень требований к технической (приборной) базе; различия требований к кадровому составу; различия в форме научных результатов и некоторые другие.

Процедуры балльного оценивания типовых показателей (включая группировку организаций по принципу сопоставимости показателей) мало зависят от специфики конкретной задачи кластеризации и потому могут рассматриваться как универсальные.

Полученные балльные оценки определяют кластеры научных организаций, образованные по принципу «каждое сочетание значений оценок есть отдельный кластер».

На четвертом этапе, в целях сокращения числа рассматриваемых кластеров, могут быть построены сводные оценки показателей. Предполагается, что каждая свертываемая группа оценок R1, R2 ... Rm может быть заменена одной сводной оценкой R с помощью некоторой функции F (функции свертки):

R = F (R1, W1, R2, W2 ... Rm, Wm) (1),

где Wn — вес оценки Rn, n = 1, 2 … m; веса здесь — дробные числа, выбираемые в соответствии с рекомендациями экспертов.

Обычной логике сверток соответствует линейная функция:

R = R1*W1 + R2*W2 + ... + Rm*Wm (2).

Для получения целочисленных значений R после вычислений по формуле (2) необходимо применить округление.

При необходимости, могут строиться достаточно сложные нелинейные соотношения, определяющие свертки вида (1).

Построение сводных оценок удобно начинать с формирования очевидных по смыслу сверток небольших групп однотипных по смыслу оценок. Примерами могут служить названные ниже оценки.

1. Оценка образовательной деятельности (обобщает отдельные оценки численности студентов, аспирантов и соискателей, проходящих обучение либо стажировку в организации).
2. Оценка кадрового потенциала (объединяет оценки численности разных категорий работников, а также оценки текучести научных кадров).
3. Оценка материально-технического потенциала (характеризует имущественный комплекс организации).

Такие свертки легко сделать типовыми как по виду используемых функций сверток, так и по значениям весов.

От сверток небольших групп оценок удобно переходить к более крупным сверткам по базовым направлениям. Их также можно сделать типовыми по виду используемых функций сверток и по значениям весов.

Разумеется, переход к сверткам допустим лишь постольку, поскольку не ведет к излишнему укрупнению кластеров. По этой причине переход от сверток по небольшим группам оценок либо по базовым направлениям к единой окончательной свертке не всегда возможен. В тех случаях, когда

этот переход возможен, его выполнение существенно зависит от смысла искомых кластеров. Типизация здесь удается (и то не всегда) только по виду функций сверток, но веса специфичны для каждой задачи.

Окончательное решение, таким образом, может быть представлено одной или несколькими балльными оценками, носящими, возможно, характер сверток.

Отметим, что число градаций шкалы свертки может отличаться от градаций шкал свертываемых показателей. Регулирование числа градаций свертки можно осуществлять как регулированием числа градаций свертываемых оценок, так и регулированием их весов. Например, в формуле (2) сумма весов не обязательно должна выбираться равной единице. Если она превосходит 1, то шкала оценки R способна получить число градаций, превышающее число градаций свертываемых оценок.

Таким образом, процесс кластеризации научных организаций в соответствии с предлагаемой методикой сводится к следующему:

1) отбираются научные организации, подлежащие анализу в рамках конкретной задачи выделения кластеров;
2) на основе типовых списков показателей осуществляют выбор учитываемых показателей для всех рассматриваемых научных организаций (по данным статистики, бухгалтерской отчетности и др. определяются значения этих показателей);
3) переход к целочисленным балльным оценкам значений показателей с применением типовых процедур балльного оценивания;
4) с использованием типовых сверток находятся сводные оценки по небольшим группам оценок, а также по базовым направлениям кластеризации; они служат основой для построения окончательного набора оценок;
5) по окончательному набору оценок формируются искомые кластеры, различающиеся значением как минимум одной из окончательных оценок.

Не претендуя на получение абсолютных результатов, данный подход является довольно универсальным и может быть использован при решении широкого круга задач государственного регулирования научно-технической сферы.

Литература

1. Нечаева Е.Г., Орлов А.И., Соколов А.В. Статистика объектов нечисловой природы и анализ данных о научном потенциале // [Электронный ресурс] http://www.isras.ru/files/File/4M/5-6/Nechaeva,%20Orlov,%20Sokolov.pdf
2. Гиршов Е. Алгоритмы кластеризации. Семинар «Введение в Computer Science» // [Электронный ресурс] http://teormin.ifmo.ru/education/intro/
3. Гради Буч. Объектно-ориентированный анализ и проектирование / Пер. с англ. под ред. И. Романовского и Ф. Андреева // [Электронный ресурс] http://www.helloworld.ru/texts/comp/other/oop/index.htm